ISSN 1897-1652

POLISH JOURNAL OF PHILOSOPHY

Volume VII, No. 2 (2013)

Jagiellonian University

Kraków

POLISH JOURNAL OF PHILOSOPHY
FALL 2013

Discussion

Book Reviews

POLISH JOURNAL OF PHILOSOPHY
Vol. VII, No. 2 (Fall 2013), 9-20.

Three Heresies in Logic and Metaphysics

Graham Priest
Departments of Philosophy, the Graduate Center, CUNY, and the University of Melbourne

Abstract. This paper concerns three heterodox views in logic and metaphysics: dialetheism (the view that some contradictions are true), noneism (the view that some objects do not exist), and the non-transitivity of numerical identity. It explains each of the views, some of their features and applications, and some of the relationships between them.

1. Introduction: Orthodoxy and Heterodoxy

Orthodoxies are strange things. Almost by definition, if something is an orthodox belief at a time and place, most of the people there and then accept it; and this very fact has a tendency to blind people to how fragile orthodoxy is. Most orthodoxies in politics, religion, science, philosophy, started out life as heterodoxies, and are fated, in their turn, to be replaced by novel heterodoxies (which is not, of course, to say that novel heterodoxies are bound to flourish – far from it). If nothing else, one would hope that an appreciation of this fact would engender an appropriate humility with respect to the things we are wont to take for granted.

Much of my work in logic and philosophy has concerned various views that are heterodox – even heretical – by contemporary standards.[1] Whether they will ever become anything else, only time will tell. But they serve, at least, to challenge an unhealthy dogmatism.

[1] To what extent they are historically so, is a different matter. On these see Priest (2007), (2008), (2010a).

DOI: 10.5840/pjphil20082219

The point of what follows is simply to explain briefly what these views are. Three, in particular, will concern us: dialetheism, noneism, and nontransitive identity.

2. Heresy Number 1

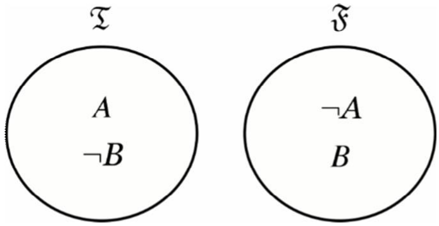
2.1. Dialetheism and Paraconsistency

The first heterodoxy (the oldest on my part, and perhaps the most provocative of the three), is dialetheism, and its corresponding paraconsistency. Dialetheism is a metaphysical view: *some contradictions are true*. That is, where $\neg$ is negation, there are sentences, propositions (or whatever one takes truth-bearers to be), A, such that both A and $\neg A$ are true. Given that A is false iff its negation is true, this is to say that there are some As which are both true and false.

Paraconsistency is a property of a relation of logical consequence. *Explosion* is the property of such a relation according to which any contradiction implies anything. That is, a relation of logical consequence, is explosive iff for all A and B, $\{A, \neg A\} \vdash B$. A consequence relation is paraconsistent iff it is not explosive. There is, of course, a connection between dialetheism and paraconsistency. In particular, if one is a dialetheist, one had better hold that the appropriate logical consequence relation is paraconsistent, on pain of accepting everything: triviality.

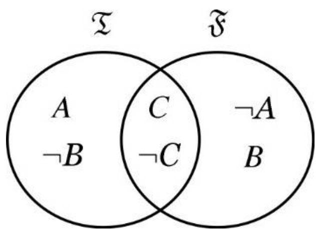
Those who have never met dialetheism and paraconsistency before may well be puzzled by the views. Let me do at least a little to unpuzzle.[2] In classical logic, any situation (interpretation) partitions all truth-bearers into two classes, the true ($\mathfrak{T}$) and the false ($\mathfrak{F}$). The two classes are mutually exclusive and exhaustive. Given that negation is a functor which toggles a sentence between truth and falsity, a sentence is in $\mathfrak{T}$ iff its negation is in $\mathfrak{F}$, and vice versa. Thus, we have the following:

[2] There is, in fact, a wide variety of paraconsistent logics. A full discussion can be found in Priest (2002). What I will describe is the basic idea if the logic is characterised model-theoretically.



A standard definition of validity is to the effect that an inference is valid if there is no situation where all the premises are true and the conclusion is not true. Given the above set-up, there is no situation where, for any A, both A and $\neg A$ are true. *A fortiori*, there is no situation where A and $\neg A$ are true, and B is not – whatever B you choose. That is, Explosion is valid.

But now suppose that $\mathfrak{T}$ and $\mathfrak{F}$ may actually overlap in some situations. Given that negation works in the same way, it follows that if C is in the overlap, so is its negation. Thus, we have the following:



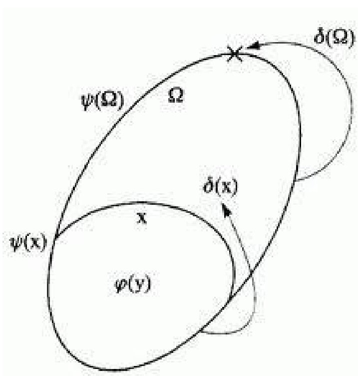
For a situation of this kind, both C and $\neg C$ are true (and false as well; but at least true). But B is false only (not true). Given exactly the same definition of validity as before, it follows that Explosion is not valid.

2.2. Dialetheism and the Inclosure Schema

Note that validity is defined as truth-preservation over *all* situations of a certain kind. These do not all have to be actual (that is, where everything that holds in the situation is actually true). One can reason correctly, not just about actual situations, but about ones that are hypothetical, conjectural, and so on. Moreover, not all situations may be expected to be such that $\mathfrak{T}$ and $\mathfrak{F}$ have a proper overlap. There will be consistent situations in which they are exclusive and exhaustive – and so where things behave *exactly* as in classical logic. So one does not have to be a dialetheist to hold that a paraconsistent logic gives the right (or a right) notion of logical consequence. One might hold that *actual* situations are of the consistent kind. Inconsistency, when it arises, occurs in merely hypothetical situations, such as the one we consider when we think what it would be like if someone squared the circle, or as described by some inconsistent (and untrue) scientific theory, such as Bohr's theory of the atom.

Dialetheism is, however, the view that some *actual* situations are such that, in them, $\mathfrak{T}$ and $\mathfrak{F}$ have a proper overlap. One may naturally, at this point, ask for examples of things in such an overlap. There are, in fact, many such – albeit controversial – examples, concerning change, boundaries, norms;[3] but perhaps the most frequently cited examples are those provided by the paradoxes of self-reference. Let us look more closely at these.

An *inclosure-paradox* is a paradox that fits the inclosure schema. This arises when there is an operator, δ, and a totality, Ω, such that whenever δ is applied to any subset, x, of Ω, of a certain kind – that is, one which satisfies some condition ψ – it appears to deliver an object that is still in Ω, though it is not in x. A contradiction will then arise if Ω itself satisfies ψ. For applying δ to Ω itself will produce an object that is both within and without Ω. We may depict the situation as follows. (φ is the defining condition of the set Ω, and $\times$ marks the contradictory spot – somewhere that is both within and without Ω):



Now, all the standard paradoxes of self-reference are inclosure paradoxes.[4] Consider, for example, Russell's paradox. If x is any set, then the set of all its members which are not members of themselves is a set, and can easily be shown not to be a member of x. Applying this operation to the totality of all

[3] See Priest (1987).

[4] See Priest (1995), Part 3.

sets therefore gives a set that is both in and not in that totality. In this case, Ω is the set of all sets; $\delta(x) = \{y \in x : y \notin y\}$; and $\psi(x)$ is the vacuous condition, $x = x$.

Some of the paradoxes in question are paradoxes of definability. A paradigm of these is König's paradox. Something is *definable* if there is a (nonindexical) noun-phrase that refers to it. If a is a definable set of definable ordinals, then (since this is countable), there is a *least ordinal greater than all the members of a*. It is obviously not a member of a, but it is definable by the italicised phrase. Since the set of all definable ordinals is itself definable, we may apply this operator to it to obtain an ordinal that cannot be referred to (defined), but which yet can. In this case, Ω is the set of all definable ordinals; $\psi(x)$ is 'x is definable'; and $\delta(x)$ is the least ordinal greater than all the members of x.

So much for some of the nuts and bolts of dialetheism. There is, of course, much more to be said about it. In particular, one might think that the view cannot be right, since it queers the pitch concerning truth, rationality, communication, or some such notion. It does not.[5]

3. Heresy Number 2

3.1. Noneism, Existence, and Quantification

Let us pass on to the second heterodoxy, noneism: *some things do not exist*.[6] These include: fictional characters, such as Sherlock Holmes; failed objects of scientific postulation, such as the mooted planet Vulcan; God (any one that you don't believe in). Yet we can think of such objects, fear them, admire them, just as we can existent objects. Indeed, we may not know whether an object to which we have an intentional relation of this kind exists or not. We may even be mistaken about its existential status. The domain of objects comprises, then, both existent and non-existent objects. There is a monadic existence predicate, E, whose extension is exactly the set of existent objects; and the extension of an intentional predicate, such as 'admire,' is a set of ordered pairs, the first of which exists, and the second of which may or may not. How to understand the notion of existence, is, of course, a thorny issue. I take it to be having the potential to enter into causal relations.

---

[5] See Priest (2006).

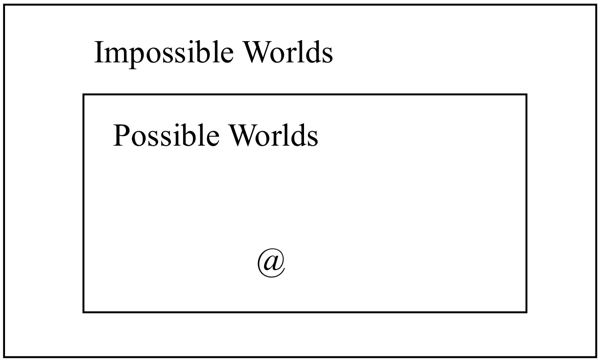
[6] This view is articulated and defended in Priest (2005).

We can also quantify over the objects in the domain, whether or not they exist. Thus, if I admire Sherlock Holmes, I admire *something*; and I might want to buy *something*, only to discover that it does not exist. I write the particular and universal quantifiers as $\mathfrak{S}$ and $\mathfrak{A}$, respectively. Normally one would write them as $\exists$ and $\forall$, but given modern logical pedagogy the temptation to read $\exists$ as 'there exists' is just too strong. Better to change the symbol for the particular quantifier (and let the universal quantifier go along for the ride). Thus, one should read $\mathfrak{S}xPx$ as 'some x is such that Px' (and $\mathfrak{A}xPx$ as 'all x are such that Px'). It is not to be read as 'there *exists* an x such that Px' – or even as 'there *is* an x such that Px', being (in this sense) and existence coming to the same thing. To put it in Meinongian terms, some objects have *Nichtsein* – non-being. If one wants to say that there *exists* something that is P, one needs to use the existence predicate explicitly, thus: $\mathfrak{S}x(Ex \wedge Px)$. Quantifiers, note, work in the absolutely standard fashion: $\mathfrak{S}xPx$ is true iff something in the domain of quantification satisfies Px; and $\mathfrak{A}xPx$ is true iff everything in the domain of quantification satisfies Px.

3.2. Noneism and Characterisation

So far so good. But more needs to be said about the properties of non-existent objects. Consider the first woman to land on the Moon in the 20th century. Was this a woman; did they land on the Moon? A natural answer is yes: an object, characterised in a certain way, has those properties it is characterised as having (the Characterisation Principle). That way, however, lies triviality, since one can characterise an object in any way one likes. In particular, we can characterise an object, a, by the condition that $x = x \wedge A$, where A is arbitrary. Applying the Characterisation Principle gives $a = a \wedge A$, and A follows. We must take a different tack.

Worlds are many. Some of them are possible; some of them are impossible. The actual world, @, is one of the possible ones:



According to dialetheism, there are contradictions true at the actual world. One might wonder, therefore, what makes a world impossible. Answer: an impossible world is one where the laws of logic are different from those of the actual world (in the way that a physically impossible world is a world where the laws of physics are different from those of the actual world).

Given the plurality of worlds, truth, truth conditions, and so on, must be relativised to each of these. That is a relatively routine matter. What is not so routine is Characterisation. If we characterise an object in a certain way, it does indeed have the properties it is characterised as having; not necessarily at the actual world, but at some world (maybe impossible). Specifically, suppose we characterise an object as one satisfying a certain condition, Px. We can write this using an indefinite description operator, ε, so that εxPx is 'an x such that Px'.[7] Given that we play our paraconsistent cards right, for *any* condition, Px, this is going to be satisfied at *some* worlds. If @ is one such, the description denotes an object that satisfies the condition there. If not, just take some other world where it is satisfied, and some object that satisfies it there. The description denotes that. Hence, we know that if $\exists xPx$ is true at @, so is $P(\varepsilon xPx)$; but if not, $P(\varepsilon xPx)$ is true at least at some world. Thus, consider the description εx(x is the first woman to land on the Moon in the 20th century). Let us use 'Selene' as a shorthand for this. Then we can think about Selene, realise that Selene is non-existent, etc. Moreover, Selene does indeed have the properties of being female and of landing on the Moon – but not at the actual world. (No existent woman was on the Moon in the 20th century; and no non-existent woman either: to be on the Moon is to interact causally with its surface, and therefore to

[7] To obtain a definite description, we just have to use the indefinite description 'an object uniquely satisfying Px'.

exist.) Selene has those properties at a (presumably possible) world where NASA decided to put a woman on one of its Moon flights.

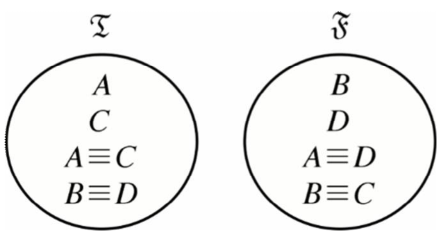
4. Heresy Number3

4. 1. Non-Transitive Identity and Leibniz' Law.

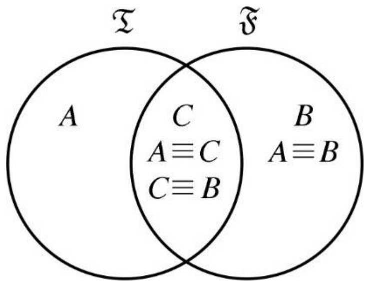
Let us now move on to the third heterodoxy (and the newest on my part): (*numerical*) *identity is not transitive*. That is, the inference $a = b,\ b = c \vdash a = c$ is not valid.

Identity, I take it, is to be defined in a familiar way, by Leibniz' Law. Two objects are the same if one object has a property just if the other does. In the language of second-order logic, $a = b$ iff $\mathfrak{A}X(Xa \equiv Xb)$. The second-order quantifiers here are to be taken as ranging over all properties – whatever, exactly, those are – and behave in a fairly standard way. The material biconditional is, however, that of a paraconsistent logic. This makes a big difference.

Classically, as I noted, every situation partitions sentences of the language into two zones, the truths ($\mathfrak{T}$) and the falsehoods ($\mathfrak{F}$), the two zones being mutually exclusive and exhaustive:



Sentences, *A*, *B*, *C*, ... therefore find themselves in exactly one or other of the zones. If two sentences are both in the same zone, their material equivalence is in the $\mathfrak{T}$ zone; whilst if one is in one zone, and the other is in the other zone, their material equivalence is in the $\mathfrak{F}$ zone. (See the diagram above.) In paraconsistent logic, everything is the same except that the $\mathfrak{T}$ and the $\mathfrak{F}$ zones may overlap. Thus we have the following picture:



As before, the material equivalence of two sentences is in the $\mathfrak{T}$ zone if both are in the same zone ($\mathfrak{T}$ or $\mathfrak{F}$), and in the $\mathfrak{F}$ zone if they are in different zones, but now a sentence can be in both zones.

$A \equiv A$ will always be in the $\mathfrak{T}$ zone, since A is always in the same zone as itself. If $A \equiv B$ is in the $\mathfrak{T}$ zone, then so is $B \equiv A$, since these are just ways of saying that A and B are in the same zone. So equivalence is reflexive and symmetric; but it is not transitive. A and C may be in the same zone, and C and B may be in the same zone, though A and B are not, because C is in the overlap. Hence, we may have $A \equiv C$ and $C \equiv B$ being in the $\mathfrak{T}$ zone, without $A \equiv B$ being so (see the above diagram). Note also that detachment for $\equiv$ may fail: we can have C and $C \equiv B$ in the $\mathfrak{T}$ zone without B being in it (same diagram).

Remembering that validity is defined as truth preservation over all interpretations, it is not difficult to check the validity of the following inferences:

- $A, B \vdash A \equiv B$
- $\neg A, \neg B \vdash A \equiv B$
- $A, \neg B \vdash \neg(A \equiv B)$
- $A, B \vdash \neg A \equiv \neg B$

Now, identity is going to inherit its behaviour from that of the biconditional. In particular, it is going to be reflexive, $\vdash a = a$, and symmetric, $a = b \vdash b = a$, but not transitive. Suppose, for the sake of illustration, that there is only one property in question, P, and that Pa, Pb and $\neg Pb$, and $\neg Pc$. Then $Pa \equiv Pb$, $Pb \equiv Pc$, but not $Pa \equiv Pc$ (since Pa and Pc are not together in either the $\mathfrak{T}$ zone or the $\mathfrak{F}$ zone). Since P is the only property at issue, we have $a = b$ and $b = c$, but not $a = c$.

Of course, the fact that the transitivity of identity does not always preserve truth does not mean that it *never* does. It is not difficult to check that:

- $A \equiv B, B \equiv C \vdash (A \equiv C) \vee (B \wedge \neg B)$

(where conjunction and disjunction work in the usual way). Hence, the transitivity of the biconditional will fail when the medial formula is both true and false, but only then. Generalising:

- $\mathfrak{A}X(Xa \equiv Xb),\ \mathfrak{A}X(Xb \equiv Xc) \vdash \mathfrak{A}X(Xa \equiv Xc) \vee \mathfrak{S}X(Xb \wedge \neg Xb)$

That is:

- $a = b,\ b = c \vdash a = c \vee \mathfrak{S}X(Xb \wedge \neg Xb)$

Hence, the transitivity of identity will fail when the medial object has a contradictory property, but only then.

4.2. Vagueness and Unity

So where might the failure of transitivity of identity be philosophically important? One obvious example concerns certain sorts of sorites paradoxes.[8] Suppose that we have a series of colour patches, $a_0, \ldots, a_n$, such that successive patches are phenomenologically indistinguishable; but that a_0 is clearly red, and a_n is clearly blue (not red). Then we have: $a_0 = a_1$, $a_1 = a_2$, ..., $a_{n-1} = a_n$. But it is not the case that $a_0 = a_n$. So transitivity fails. Should we expect this? Yes. This sorites paradox can be seen as an inclosure paradox.[9] $\Omega = \{a_i : a_i \text{ is red}\}$. $\psi(x)$ is the vacuous condition, $x = x$. If $x \subseteq \Omega$, then there is a greatest i such that $a_i \in x$; $\delta(x) = a_{i+1}$. Clearly, $\delta(x) \notin x$, but since a_{i+1} is indistinguishable from a_i, $\delta(x)$ is red, and so $\delta(x) \in \Omega$. Hence, a medial object, $\delta(\Omega)$, has contradictory properties: it is both red and not red.

A much less obvious example of the application of non-transitive identity is the following.[10] Consider any object (existent or non-existent), x, with parts a, b, c, ... There is a difference between the object, x, which is one, and the plurality of its parts. Something must bind the parts into a unity. Call this the *gluon* of x, g. Is g, or is it not, itself, an object? It is an

[8] On what follows, see Priest (2010b).
[9] See Priest (2010c).
[10] On the following, see Priest (2014).

object, since we can think about it, refer to it, quantify over it. But it is not an object. If it were, *g*, *a*, *b*, *c*, ... would be just as much a congeries as the original plurality of parts. For if the gluon were an object which joins the parts together, there must be something which joins *g* to each of *a*, *b*, *c*, ...; and we are off on a vicious regress. (The Bradley Regress.) A gluon therefore has contradictory properties: it both is and is not an object.

But, inconsistent though it be, how does it manage to make the parts a whole? Simply by being identical to each of the other parts (and itself). Thus, $g = a$, $g = b$, $g = c$, ... In this way is the Bradley Regress broken. Since $g = a$, there is no metaphysical space – as it were – or need, for anything to be inserted between them to join them;[11] and the same is true for *b*, *c*, *d*, ... So we have $a = g = b$; but of course, it is not the case that $a = b$. Hence, identity is non-transitive. Moreover, we know that transitivity should be expected to fail exactly when the medial object has contradictory properties. Since this is a gluon, this is, indeed, the case.

5. Conclusion: *Being*

So for our three heresies. Some may think that only ill can come from compounding heresy upon heresy. Personally, I do not see it that way. The orthodoxies on these matters were never as rationally grounded as their adherents like to pretend. Moreover, the three heresies, far from adding to each others' woes, interlock and support each other in fundamental ways, as should be clear.

Let me give one final example of their interaction. Heidegger famously asked the *Seinsfrage*. What is it to be? What is it in virtue of which a being is? By being, here, he most certainly does not mean existence (*existenz*). By *being*, he means *being an object*. (In Meinongian terms, the sense of *being* is *Außersein*.) Heidegger gave up trying to answer the question explicitly, since doing so drives one into contradiction. *Being* is not itself a being; but any answer of the form '*being* is such and such' treats it as exactly that.[12]

So what is it to be an object? As Aristotle noted a long time ago, to be is to be one. So what makes a thing (to) be is what makes it (to be) one. That is its gluon. And since this both is and is not an object, one can and cannot refer to it. (One can speak about objects and only objects.) In the same way, one can and cannot refer to the least indefinable ordinal. We have, then, an answer to the *Seinsfrage*; and one, moreover, that precisely explains

[11] And what joins *a* to *g*? *g* itself, since $a{=}g$ and $g{=}g$.

[12] See the second edition of Priest (1995), ch. 15.

Heidegger's predicament. Perhaps the possibility of such an answer is Heresy Number 3½. Thus can our three heresies combine to open up an entirely new perspective on the world, and one forever closed to those with the blinkers of current orthodoxy.

References

Priest, G. (1987). *In Contradiction*. Dordrecht: Martinus Nijhoff; second (extended) edition, Oxford: Oxford University Press, 2006.

Priest, G. (1995). *Beyond the Limits of Thought*. Cambridge: Cambridge University Press; second (extended) edition, Oxford: Oxford University press, 2002.

Priest, G. (2002). Paraconsistent Logic. In Gabbay, D., Guenthner, F. (eds.). *Handbook of Philosophical Logic*. *Volume 6* (pp. 287-393). 2nd edition, Dordrecht: Reidel.

Priest, G. (2005). *Towards Non-Being*. Oxford: Oxford University Press.

Priest, G. (2006). *Doubt Truth to be a Liar*. Oxford: Oxford University Press.

Priest, G. (2007). Paraconsistency and Dialetheism. In Gabbay, D., Woods, J. (eds.). *Handbook of the History of Logic*. *Volume 8* (pp. 129-204). Amsterdam: North Holland.

Priest, G. (2008). The Closing of the Mind: How the Particular Quantifier Became Existentially Loaded Behind our Backs. *Review of Symbolic Logic*, *1*, 42-55.

Priest, G. (2010a). A Case of Mistaken Identity. In Lear, J., Oliver, A. (eds.). *The Force of Argument* (pp. 205-222). London: Routledge.

Priest, G. (2010b). Non-Transitive Identity. In Dietz, R.J. Moruzzi, S. (eds.). *Cuts and Clouds: Vagueness, its Nature and its Logic* (pp. 406-416). Oxford: Oxford University Press.

Priest, G. (2010c). Inclosures, Vagueness, and Self-Reference. *Notre Dame Journal of Formal Logic*, *51*, 69-84.

Priest, G. (2014). *One*. Oxford: Oxford University Press.